A neuroeconomic theory of (dis)honesty *

Isabelle Brocas
University of Southern California and CEPR

Juan D. Carrillo
University of Southern California and CEPR

July 2017

Abstract

We develop a theory of dishonesty based on neurophysiological evidence that supports the idea of a two-step process to decide whether to cheat. The individual first decides on a rule of conduct by choosing whether to be open to accept bribes or not. Being open to bribes is costly but it gives the option to agree to be bribed based on the specific stakes of each event. The model has two testable predictions that depart from traditional theories. First, controlling for the size of the bribe, expectations about bribes affect the likelihood of dishonest behavior. Consistent with the idea that cheating is driven by a rule of conduct, an individual will accept a given bribe more often in an environment in which bribes are high (in which he would choose to be open to bribes) than in an environment in which bribes are low (in which he would commit to never accept them). Second, this expectation effect results in multiple self-fulfilling equilibria in the market. The latter exhibits either high bribes and a high level of dishonesty (bribees are likely to be open to bribes) or low bribes and a low level of dishonesty (bribees are likely to never accept bribes).

Keywords: neuroeconomic theory, control network, dishonesty, bribery.

JEL Classification: D73, D87.

---

*Address for correspondence: Department of Economics, University of Southern California, 3620 S. Vermont Ave., Los Angeles, CA 90089, USA, emails: `<brocas@usc.edu>` and `<juandc@usc.edu>`.

We are grateful to members of the Los Angeles Behavioral Economics Laboratory (LABEL) and the audience at various seminars for comments and suggestions.

1 Introduction

The economics of corruption, dishonesty and collusion has been the subject of significant research in economics. From the classical analyses of dishonesty in politics, bureaucracies and other institutions (Ackerman, 1978; Klitgaard, 1988) to the formal models of corruption in firms (Lui, 1986; Tirole, 1996). Corruption has been studied from experimental (Rosenbaum et al., 2014; Abeler et al., 2016), empirical (Fisman and Gatti, 2002; Fisman and Miguel, 2007), and theoretical (Tirole, 1992; Kofman and Lawarree, 1993) perspectives. More recently, it has been the object of studies in behavioral economics (Dal Bó and Terviö, 2013) and neuroscience (Baumgartner et al., 2009; Greene and Paxton, 2009).

Despite important advances, our knowledge of the factors conducive to dishonest behavior is still incomplete. In this paper, we incorporate recent experimental evidence from the brain sciences to build a novel theory of dishonesty. Our goal is not to depart from the traditional neoclassical approach where subjects evaluate the costs and benefits of illegal actions and act accordingly. Instead, we extend the standard setting by incorporating neurophysiological considerations and obtain a larger set of behavioral predictions.

The basic premise of our theory is that cheating involves a two-step process. First, the individual decides whether to act honestly independently of the situation or to consider the possibility of being dishonest. The former is costless whereas the latter entails a psychological cost in terms of attentional or cognitive resources necessary to evaluate the different options. Second, the individual who opens the door to dishonesty trades-off the monetary and non-monetary costs and benefits of cheating and acts accordingly.

Using this framework, we first show that increasing the attentional and cognitive demands –for example, by requiring subjects to engage in multi-tasking– makes dishonesty relatively more costly and therefore less prevalent (Proposition 1). Perhaps more subtle, we also show that the distribution of rewards derived from the illegal services affects the decision to behave honestly. If rewards for illegal services in the economy are higher, it is more interesting to pay the cost of being open to dishonesty. Once this cost is sunk, dishonest behavior is more likely to follow (Proposition 2). Finally, the effect of expected rewards on dishonesty is self-fulfilling and results in multiple equilibria with different levels of corruption in the population. High anticipated bribes imply high incentives to be open to dishonesty which itself makes it worthwhile for bribers to offer high compensations.

Conversely, low anticipated bribes imply low incentives to be open to dishonesty which makes it less valuable for bribers to offer high compensations (Proposition 3).

The paper is organized as follows. In section 2, we develop the standard model of dishonesty and our extended version based on neurophysiological evidence. In section 3, we present the comparative statics results of our new model. In section 4, we discuss the implications of the model in a bribery game. In section 5, we offer some concluding comments.

2 A model of costly dishonesty

2.1 The canonical model

In the canonical economic model of corruption (Lui, 1986; Andvig and Moene, 1990), individuals trade off the costs and benefits of engaging in an illegal, immoral, or otherwise reprehensible activity (from now on, we will generically refer to as "cheating"). Formally, the payoff of such activity is captured by the following utility function:

$$u = b - k\,\theta \tag{1}$$

In this equation, $b \in [\underline{b}, \overline{b}] \subset \mathbb{R}^+$ represents the net *monetary* benefit associated with cheating. It may refer for instance to a bribe received net of the expected punishment, or a payment for an illegal service. The (possibly unobserved) parameter $\theta \in [\underline{\theta}, \overline{\theta}] \subset \mathbb{R}^+$ captures the intrinsic *honesty* level of the individual, that is, the non-monetary disutility of cheating. It may represent a psychological cost (guilt, shame, aversion to lying, loss of self-respect, etc.) or an economic cost (reputation loss vis-à-vis others, trustworthiness that affects future trade possibilities, etc.). Finally, k (> 0) reflects how important monetary benefits are compared to honesty concerns.

A key characteristic of the canonical model is that individuals differ in their level of honesty θ (Tirole, 1996; Carrillo, 2000a,b). In particular, an individual whose θ is sufficiently high never cheats whereas an individual whose θ is sufficiently low always cheats.[1] We denote by $G(\theta)$ the cumulative distribution of honesty levels in the population. In this framework and for a given monetary benefit b, a fraction of individuals $G(\check{\theta})$ chooses to cheat and a fraction $1 - G(\check{\theta})$ chooses not to. Given the preferences described in (1),

[1]This is under the implicit assumption that $\overline{b} \leq k\overline{\theta}$ and $\underline{b} \geq k\underline{\theta}$ but it can be easily generalized.

we have $\check{\theta} = b/k$. Not surprisingly, the amount of cheating in the population is increasing in the size of the monetary gain ($dG(\check{\theta})/db > 0$).

2.2 A neurophysiologically based model of dishonesty

Although this basic model is an excellent first approximation and has delivered numerous insights, it does not incorporate important elements of dishonesty. Our objective is to use evidence from neuroscience to develop a more comprehensive model of "costly dishonesty" in order to provide new implications. The model relies on the experimental findings obtained by Baumgartner et al. (2009) (hereafter, [B.al]) and Greene and Paxton (2009) (hereafter, [GP]) that identify *neural correlates of dishonesty.* We will model [GP], as it focuses primarily on the choice between honest behavior and dishonest behavior. In the [GP] experiment, subjects are asked to predict the outcome of computerized coin flips. In the baseline condition, they record their predictions in advance and are compensated based on accuracy. In the treatment condition, predictions are not recorded and rewards are based on self-reported accuracy. Subjects are behaviorally classified as "honest", "dishonest" or "ambiguous" depending on whether the self-reported success rate in the 70 trials of the treatment condition is average ($\leq .59$), improbably high ($\geq .69$) or in-between. Interestingly, dishonest individuals show evidence of cheating but not in every trial. A comparison of neural activity between the baseline and treatment conditions reveals two important findings. First, there is no significant difference in activity between the baseline and treatment conditions for subjects categorized as honest. Second, there is increased activity in the so-called "control network" (anterior cingulate, dorsolateral prefrontal cortex and ventrolateral prefrontal cortex) in the treatment condition for subjects categorized as dishonest. This extra activity occurs both for self-reported successes (which pools correct predictions and truthful reports with incorrect predictions and false reports) and for self-reported failures (cases unambiguously characterized by incorrect predictions and truthful reports). The experiment in [B.al] is slightly different but yields very similar qualitative results and reveals similar patterns of activation.[2]

The [GP] experiment suggests the following two-step decision process. Before the

[2]More precisely, [B.al] also studies neural correlates before the honesty choice stage (when subjects decide whether to make a promise that they will later break or not, and when subjects observe the reaction of others to the promise they plan to break or not). Dishonest subjects break their promise of sending back half the money around 70% of the time. They also show activity when breaking the promise in remarkably similar areas (anterior cingulate, dorsolateral prefrontal cortex and amygdala).

experiment starts, the subject knows the distribution of benefits $F(b)$ but not the benefit in each trial. She also knows the ex ante probability α that her true prediction (though not necessarily the reported one) is incorrect.[3] She decides once-and-for-all (step 1) whether to be honest (i.e., always reveal truthfully her prediction) or to be dishonest (i.e., *selectively* choose when to reveal truthfully) as a function of her honesty level θ. The decision to be dishonest engages the control network, independently of the final choice. Recruiting the control network involves a cost c that reflects the attentional or cognitive resources necessary to evaluate the options. This feature captures the above mentioned neuroscience evidence: it is the once-and-for-all decision to entertain the possibility of cheating that engages the control network (and not the cheating itself).

The first step decision is then implemented in each trial (step 2). If the subject has decided to be honest, she provides a truthful report in each trial. If, instead, she has decided to be dishonest, she chooses on each trial whether to tell the truth (no cheating) or to lie (cheating) depending on the accuracy of her prediction and the stakes in that trial. The monetary benefit b is obtained when the self-report (whether truthful or not) coincides with the outcome and the dishonesty cost θ is incurred when the self-report is not truthful. This process is summarized in Figure 1.

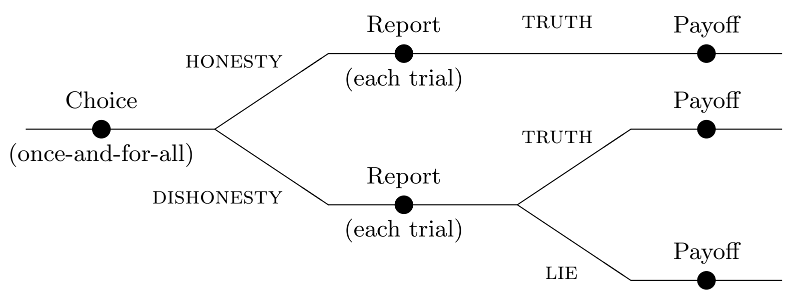

Figure 1. Timing

Assuming for simplicity that $k = 1$, the second step utility of a subject depends on her choice to be honest or not in the first step. Formally, if the subject has decided to be honest and to always report truthfully her prediction, she obtains a payoff b when it coincides with the true outcome and she obtains no payoff otherwise. If the subject decides instead to be dishonest, two cases are possible. When her prediction is correct, she has

[3]In [GP]'s experiment, subjects predict a coin toss, so $\alpha = 1/2$. The experimenter can easily manipulate these objective probabilities (e.g., dice toss) or make them contingent on the subject's expertise.

no incentives to misreport it. When her prediction is incorrect, she will choose to cheat if and only if $b > \theta$.[4] The expected utility of choosing to be honest at the beginning of step 1 is independent of θ and it is given by:

$$u_H = (1-\alpha)\, E[b] \tag{2}$$

On the other hand, the expected utility of choosing to be dishonest at the beginning of step 1 is:

$$u_D(\theta) = (1-\alpha)\, E[b] + \alpha \int_{b=\theta}^{\bar{b}} (b-\theta) f(b) db - c \tag{3}$$

Note that a subject who chooses to be dishonest pays the cost c independently of the final choice. This feature is referred to as "costly dishonesty".

3 Analysis of the costly dishonesty model

Comparing (2) and (3), it is immediate that the subject will choose to be dishonest if and only if:

$$u_D(\theta) - u_H \geq 0 \quad \Leftrightarrow \quad V(\theta) \equiv \alpha \int_{b=\theta}^{\bar{b}} (b-\theta) f(b) db - c \geq 0 \tag{4}$$

To be in the interesting scenario, assume that the expected benefit is sufficiently high: $E[b] > \underline{b} + c/\alpha$ so that $V(\underline{b}) > 0$. Since $V(\bar{b}) = -c < 0$ and $V'(\theta) = -\alpha[1 - F(\theta)] < 0$, it is immediate that there exists an interior cutoff $\theta^* \in (\underline{b}, \bar{b})$ given by:

$$V(\theta^*) = 0 \quad \Leftrightarrow \quad \int_{b=\theta^*}^{\bar{b}} (b-\theta^*) f(b) db = c/\alpha \tag{5}$$

such that:

$$\begin{cases} u_H > u_D(\theta) & \text{if} \quad \theta > \theta^* \\ u_H < u_D(\theta) & \text{if} \quad \theta < \theta^* \end{cases} \tag{6}$$

In words, a subject chooses to be honest if $\theta > \theta^*$ and to be dishonest if $\theta \leq \theta^*$. Then, conditional on being dishonest and making an incorrect prediction, the subject will choose to cheat if the benefit for the particular trial is $b > \theta$ and not to cheat otherwise. Our first result is the following.

[4] This deterministic rule differs from what we observe in practice both in [GP] and in [B.al], where trials with identical payoffs elicit different behaviors. Adding some uncertainty, randomness or unobserved considerations would account for such differences. The key property (which is supported by the behavioral results) is that, other things being equal, lying increases with the stakes of the trial.

Proposition 1 *If c, the cost of engaging the control network, or if $1-\alpha$, the probability of succeeding by chance, increase then: (i) fewer subjects choose to be dishonest and (ii) the total amount of cheating decreases, but (iii) the subjects who choose to be dishonest cheat more often.*

Proof: Differentiating $V(\theta^*) = 0$ we get:

$$\frac{d\theta^*}{dc} = -\frac{\partial V/\partial c}{\partial V/\partial \theta^*} = -\frac{1}{\alpha[1-F(\theta^*)]} < 0.$$

By definition, the proportion of dishonest subjects is $G(\theta^*)$ so:

$$\frac{dG(\theta^*)}{dc} = g(\theta^*)\frac{\partial \theta^*}{\partial c} < 0.$$

The total amount of cheating $J(\theta^*)$ is given by:

$$J(\theta^*) = \int_{\theta=\underline{\theta}}^{\theta^*}\int_{b=\theta}^{\bar{b}} f(b)g(\theta)db\,d\theta \;\Leftrightarrow\; J(\theta^*) = \int_{\theta=\underline{\theta}}^{\theta^*} g(\theta)[1-F(\theta)]d\theta$$

Differentiating with respect to c, we get:

$$\frac{dJ(\theta^*)}{dc} = g(\theta^*)[1-F(\theta^*)]\frac{d\theta^*}{dc} < 0.$$

Finally, the amount of cheating among the dishonest subjects $D(\theta^*)$ is:

$$D(\theta^*) = \int_{\theta=\underline{\theta}}^{\theta^*}\int_{b=\theta}^{\bar{b}} f(b)\frac{g(\theta)}{G(\theta^*)}db\,d\theta \;\Leftrightarrow\; D(\theta^*) = \frac{1}{G(\theta^*)}\int_{\theta=\underline{\theta}}^{\theta^*} g(\theta)[1-F(\theta)]d\theta$$

Again differentiating with respect to c, we get:

$$\frac{dD(\theta^*)}{dc} = -\left[\frac{g(\theta^*)}{G(\theta^*)^2}\int_{\theta=\underline{\theta}}^{\theta^*} G(\theta)f(\theta)d\theta\right]\frac{d\theta^*}{dc} > 0.$$

The differentiation with respect to α is analogous. □

The result is intuitive. If engaging the control network becomes more costly, then dishonesty becomes a less attractive option. As a result, there is a larger set of types who prefer to avoid trading off cost and benefits in each trial and, instead, choose to be always honest. This decreases the total amount of cheating. Interestingly, cheating increases among the subjects who decide to be dishonest. The reason is simply that an increase in

c makes the marginal types switch from dishonesty to honesty. But these subjects were the least likely to cheat before the increase of c. Therefore, only dishonest subjects who are very likely to cheat remain dishonest. Hence, the average amount of cheating by the dishonest subjects is increased. The argument regarding α is analogous. If success is more likely to occur by pure luck, subjects have less incentives to incur the cost of becoming dishonest.

Proposition 1 provides a testable implication of the theory. Indeed, c can be manipulated by the experimenter. For example, if subjects were asked to simultaneously perform other activities that also engage the control network, this would presumably increase the cost of trading off truthful revelation vs. lying. We should then observe a decrease in total cheating and, at the same time, an increase in average cheating by subjects who decide to be dishonest. Instead, a "traditional" theory of dishonesty would predict no relationship between dishonesty and other orthogonal activities.[5]

This first conclusion is interesting and novel. However, the prediction does not rely on the process being a two-step process. Indeed, the comparative statics would be identical if we assumed instead a one-step process with two costs, θ and c. As such, our first result cannot be used as a direct test of our model. To provide further testable implications, suppose now that the monetary benefit of cheating can be drawn from one of two different distributions, $F_1(b)$ and $F_2(b)$, such that the latter first-order stochastically dominates the former: $F_2(b) < F_1(b)$ for all $b \in (\underline{b}, \overline{b})$. In words, prizes are (stochastically) higher under F_2 than under F_1. Denote by θ_i^* the dishonesty cutoff when benefits are drawn from distribution $F_i(b)$. Using (5), these cutoffs are such that:

$$\int_{b=\theta_1^*}^{\overline{b}} (b-\theta_1^*) f_1(b) db = \int_{b=\theta_2^*}^{\overline{b}} (b-\theta_2^*) f_2(b) db$$

$$\Leftrightarrow \quad \int_{b=\theta_1^*}^{\overline{b}} \Big[1-F_1(b)\Big] db = \int_{b=\theta_2^*}^{\overline{b}} \Big[1-F_2(b)\Big] db$$

Stochastic dominance immediately implies that $\theta_2^* > \theta_1^*$. In fact, the result is related to Proposition 1. A (stochastic) increase in benefit has the same effect as a (deterministic) decrease in cost: it makes dishonesty a relatively more interesting option. More interestingly, if we consider one particular trial, we obtain the following result.

[5]Notice that such traditional theory would predict less cheating as the probability of succeeding by chance increases. However, it would not predict a higher average level of cheating among those who choose to be dishonest.

Proposition 2 *For a given benefit b, the likelihood of cheating is (weakly) higher if that benefit is drawn from $F_2(b)$ than if it is drawn from $F_1(b)$.*

Proof: Fix b. If it is drawn from F_i, the subject cheats if and only if $\theta < \min\{\theta_i^*, b\}$. Since $\theta_2^* > \theta_1^*$, then for all $b > \theta_1^*$ more cheating occurs under F_2 than under F_1. $\square$

Given our two-step process, subjects decide between honesty and dishonesty as a function of the distribution of benefits. Then, the final choice between cheating and not cheating depends on the realized benefit. If benefits are likely to be high, more subjects entertain the possibility of becoming dishonest. Once they have chosen that route, they are locked into a higher likelihood of cheating. In other words, potential rewards frame the mind of subjects on the issue of honesty, and this affects their choice once rewards are announced.[6]

It is easy to see that Proposition 2 holds only under a two-step process. It offers a further test. Indeed, the model predicts that, *controlling for the size of the reward offered*, we should observe (behaviorally) more cheating and (neurally) more activity in the control network when the distribution of rewards is tilted towards high values than when it is tilted towards low values. This prediction can be tested by varying the distribution from which rewards are drawn.

4 Implications for an economic model of dishonesty

So far the size of the reward associated with cheating has been exogenously given. In an economic game, however, such reward - hereafter referred to as bribe - are likely to be provided in exchange of some favor or other valuable service. In that case, the amount offered is determined endogenously, where supply meets demand. The objective of this section is to extend the previous model of costly dishonesty to the case of a bribery game in a general equilibrium context.

One important feature of illegal markets emphasized in the empirical literature (Klitgaard, 1988) is the existence of multiple equilibria. Given a set of initial conditions,

[6]Interestingly, the result is the opposite of an externality-driven dynamic argument in a standard economic model. Indeed, suppose that a subject who is caught cheating is fired and therefore foregoes future income. Other things being equal, higher future gains for cheating (bribes drawn from F_2 rather than F_1) increases the opportunity cost of being fired and therefore decreases the propensity to behave dishonestly in the first place.

markets may end up exhibiting a high or a low level of illegal activities for no obvious reasons. Temporary measures to combat illegal activities may also result in the market switching to a low level of illegal activities in the long run. These observations suggest that both high and low levels of illegal activities may be equilibria in some markets. The existing economics literature has investigated different extensions of the basic model to account for this multiplicity. All involve extra modeling pieces such as externalities between the number of corrupt individuals and the probability of catching each of them (Lui, 1986; Andvig and Moene, 1990) or dynamic considerations (Sah, 1991; Tirole, 1996; Carrillo, 2000b; Dal Bó and Terviö, 2013).[7] One objective of this extension is to assess whether our model of costly dishonesty may, alone, give rise to a coordination problem.

We consider a simplified (discrete-type) version of the model developed in sections 2 and 3. On the demand side, potential "bribees" differ in their degree of honesty. This is captured with the parameter $\theta \in \{0, \theta_l, \theta_h\}$ which is found in proportions p, q and $1-p-q$ in the population. On the supply side, potential "bribers" differ in the gain obtained if they are granted the illegal favor or service. This is modeled with the parameter $g \in \{g_l, g_h\}$ which is found in proportions $1-\mu$ and μ in the population. It captures the differences in the ability or opportunity cost of individuals to obtain the good legally. Finally and to simplify, we assume that bribers can only offer two levels of bribes for the illegal service: $b \in \{\underline{b}, \bar{b}\}$. The crucial novelty of the model is that bribes will now be endogenously chosen. To be in the interesting situation, we impose the following assumption:

$$0 < \underline{b} < \theta_l < \theta_h < \bar{b} < g_l < g_h \tag{7}$$

Given our basic utility formulation (1), it means that a fully dishonest individual ($\theta = 0$) is willing to accept any bribe, whereas the other two types of potential bribees (θ_l and θ_h) find it profitable to accept a high bribe ($\bar{b}$) but not a low bribe ($\underline{b}$). At the same time, both types of bribers get net benefits even if they pay a high bribe but, other things being equal, they obviously prefer to pay a low bribe.

Suppose now that bribers and bribees meet in pairs at random. The briber enjoys gain g for the illegal service and the bribee suffers dishonesty cost θ. These parameters however are private information. Whenever they meet, the briber makes a take-it-or-leave-it offer b for the service and the bribee accepts or rejects it.

[7] More precisely, those articles focus on the effect of the anticipation of future equilibrium bribery on current incentives to accept bribes.

Proposition 3 *Under some conditions about the parameters of the model, the bribery game has multiple equilibria: low bribes and low cheating or high bribes and high cheating.*

Proof. Notice that by (7), an individual who has paid the cost c of being dishonest will accept $\overline{b}$ and refuse $\underline{b}$ if $\theta = \theta_h$ or $\theta = \theta_l$, and he will accept both $\overline{b}$ and $\underline{b}$ if $\theta = 0$.

Suppose first that the briber offers $\overline{b}$ if $g = g_h$ and $\underline{b}$ if $g = g_l$. Given such offer, an equilibrium where the individual is honest if $\theta = \theta_h$ and dishonest if $\theta = \theta_l$ or $\theta = 0$ exists if:

$$\mu(\overline{b} - \theta_h) < c \qquad \text{and} \qquad \mu(\overline{b} - \theta_l) > c \tag{C1}$$

Given this behavior by bribees, it is indeed in the interest of g_h to offer $\overline{b}$ and in the interest of g_l to offer $\underline{b}$ if:

$$(g_h - \overline{b})(p + q) > (g_h - \underline{b})p \qquad \text{and} \qquad (g_l - \overline{b})(p + q) < (g_l - \underline{b})p \tag{C2}$$

Suppose now that both types of bribers offer $\overline{b}$. Given such offer, an equilibrium where all types of bribees are dishonest exists if:

$$\overline{b} - \theta_h > c \tag{C3}$$

Given this behavior by bribees, it is indeed in the interest of both type of bribers to offer $\overline{b}$ if:

$$g_l - \overline{b} > (g_l - \underline{b})p \tag{C4}$$

If conditions (C1)-(C2)-(C3)-(C4) are jointly satisfied, there are two equilibria: one where bribees with type θ_h decide to be honest (while all others are dishonest) and bribers with type g_l offer $\underline{b}$ (while those with type g_h offer $\overline{b}$) and another where all individuals are dishonest and all bribers offer high bribes. □

Multiplicity of equilibria is a consequence of our two-step process. Indeed, in our model, the potential bribee decides whether to be dishonest (and pay the cost c) before learning the size of the bribe offered by the briber. Once this cost is sunk, the trade-off is simply between bribe b and disutility θ. Consider an individual with a moderately high disutility of engaging in bribery (θ_h in our model). If he knows that all potential partners will offer high bribes, he will be willing to incur the dishonesty cost in anticipation of consistently large benefits from the exchange. In that case, he enters the market to provide illegal services. Conversely, if he realizes that he will be offered a high bribe only with some probability, he will be better-off saving the dishonesty cost even if it implies foregoing an important bribe with some chance. In that case, he refrains from entering the illegal market. On the supply side, an individual who enjoys a moderately low benefit from illegal services (g_l in our model) is willing to offer a high bribe only if it ensures obtaining the service for sure. If there is a risk involved anyways, he is better-off saving some money by proposing a smaller bribe. These incentives create a coordination problem and give rise to multiple equilibria.

Overall, our model proposes a new rationale for multiplicity of equilibria in a bribery game. The result is a direct consequence of the way decision-making operates in the context of cheating. It involves the ability of individuals to commit to some rule of conduct (never cheat or be open to cheating) and to implement those rules in their day-to-day lives. High-bribe markets correspond to situations in which bribees have committed to be open to bribery anticipating bribes will be high, and bribers offer substantial bribes understanding that bribees will consider them. Low-bribe markets correspond to the reverse situation in which bribees have committed to never cheat anticipating bribes will be low, and bribers offer low bribes understanding that some bribees will never consider them.

5 Discussion

Interdisciplinary research between neuroscience and economics has received tremendous attention in recent years, leading to the development of a new field of study: neuroeconomics. However, some economists claim that neuroscience has little to add to our knowledge of economic decision-making (Gul and Pesendorfer, 2008). While some authors have argued that understanding neural mechanisms can help make new behavioral predictions (Camerer, 2007) others insist that neuroeconomics will be useful only when it provides out-of-sample predictions in contexts of importance for economists (Bernheim, 2009).

The debate, however, has almost exclusively centered on the potential value of experimental neuroeconomics. This paper falls in a parallel agenda in neuroeconomic theory (Brocas and Carrillo, 2008; Alonso et al., 2013). It contributes to the above mentioned discussion by demonstrating the methodological advantages for the analysis of individual decision-making of combining the existing empirical fMRI evidence from neuroscience with the theoretical optimization tools of economics. Indeed, we start by arguing that traditional economic analyses of dishonest behavior are valuable but incomplete. We then show how existing experiments in neuroscience can help build a more comprehensive optimization model of cheating. The model has distinctive behavioral and neural implications that can be tested in the laboratory. These new tests may help determine if the model captures the fundamental elements of dishonesty and also provide guidelines to refine the theory, restarting the cycle theoretical model / empirical test.

For the specific issue of dishonesty, our model delivers three testable predictions that depart from traditional analyses. First, multi-tasking may overload the control network and decrease the propensity to engage in dishonest behavior. Second and controlling for the size of current rewards, higher expectations about future rewards frame the subjects' mind towards dishonesty which increases the likelihood of cheating. Finally, we point to the existence of multiple bribery equilibria, where the anticipation of the behavior of potential bribers has a self-fulfilling effect on the decision to consider the option of cheating.

References

Abeler, Johannes, Daniele Nosenzo, and Collin Raymond, "Preferences for truth-telling," 2016. Working Paper.

Ackerman, Susan Rose, "Corruption: a study in political economy," *New York: Academic Pres*, 1978.

Alonso, Ricardo, Isabelle Brocas, and Juan D Carrillo, "Resource allocation in the brain," *Review of Economic Studies*, 2013, *81* (2), 501–534.

Andvig, Jens Chr and Karl Ove Moene, "How corruption may corrupt," *Journal of Economic Behavior & Organization*, 1990, *13* (1), 63–76.

Baumgartner, Thomas, Urs Fischbacher, Anja Feierabend, Kai Lutz, and Ernst Fehr, "The neural circuitry of a broken promise," *Neuron*, 2009, *64* (5), 756–770.

Bernheim, B Douglas, "On the potential of neuroeconomics: A critical (but hopeful) appraisal," *American Economic Journal: Microeconomics*, 2009, *1* (2), 1–41.

Bó, Ernesto Dal and Marko Terviö, "Self-esteem, moral capital, and wrongdoing," *Journal of the European Economic Association*, 2013, *11* (3), 599–633.

Brocas, Isabelle and Juan D Carrillo, "The brain as a hierarchical organization," *American Economic Review*, 2008, *98* (4), 1312–1346.

Camerer, Colin F, "Neuroeconomics: using neuroscience to make economic predictions," *Economic Journal*, 2007, *117* (519).

Carrillo, Juan D, "Corruption in hierarchies," *Annales d'Economie et de Statistique*, 2000, pp. 37–61.

_ , "Graft, bribes, and the practice of corruption," *Journal of Economics & Management Strategy*, 2000, *9* (3), 257–286.

Fisman, Raymond and Edward Miguel, "Corruption, norms, and legal enforcement: Evidence from diplomatic parking tickets," *Journal of Political Economy*, 2007, *115* (6), 1020–1048.

— **and Roberta Gatti**, "Decentralization and corruption: evidence across countries," *Journal of Public Economics*, 2002, *83* (3), 325–345.

Greene, Joshua D and Joseph M Paxton, "Patterns of neural activity associated with honest and dishonest moral decisions," *Proceedings of the National Academy of Sciences*, 2009, *106* (30), 12506–12511.

Gul, Faruk and Wolfgang Pesendorfer, "The case for mindless economics," *The foundations of positive and normative economics: A handbook*, 2008, *1*, 3–42.

Klitgaard, Robert, *Controlling corruption*, Univ of California Press, 1988.

Kofman, Fred and Jacques Lawarree, "Collusion in hierarchical agency," *Econometrica*, 1993, pp. 629–656.

Lui, Francis T, "A dynamic model of corruption deterrence," *Journal of public Economics*, 1986, *31* (2), 215–236.

Rosenbaum, Stephen Mark, Stephan Billinger, and Nils Stieglitz, "Let's be honest: A review of experimental evidence of honesty and truth-telling," *Journal of Economic Psychology*, 2014, *45*, 181–196.

Sah, Raaj K, "Social osmosis and patterns of crime," *Journal of Political Economy*, 1991, *99* (6), 1272–1295.

Tirole, Jean, "Collusion and the theory of organizations," in J.-J. Laffont, ed., *Advances in Economic Theory*, Cambridge University Press 1992.

— , "A theory of collective reputations (with applications to the persistence of corruption and to firm quality)," *The Review of Economic Studies*, 1996, *63* (1), 1–22.